

The Role of Face Parts in Gender Recognition

Yasmina Andreu Cabedo and Ramón A. Mollineda Cárdenas

Dept. Llenguatges i Sistemes Informàtics
Universitat Jaume I. Castelló de la Plana, Spain.
yandreu@uji.es, mollined@uji.es

Abstract. This paper evaluates the discriminant capabilities of face parts in gender recognition. Given the image of a face, a number of subimages containing the eyes, nose, mouth, chin, right eye, internal face (eyes, nose, mouth, chin), external face (hair, ears, contour) and the full face are extracted and represented as appearance-based data vectors. A greater number of face parts from two databases of face images (instead of only one) were considered with respect to previous related works, along with several classification rules. Experiments proved that single face parts offer enough information to allow discrimination between genders with recognition rates that can reach 86%, while classifiers based on the joint contribution of internal parts can achieve rates above 90%. The best result using the full face was similar to those reported in general papers of gender recognition (>95%). A high degree of correlation was found among classifiers as regards their capacity to measure the relevance of face parts, but results were strongly dependent on the composition of the database. Finally, an evaluation of the complementarity between discriminant information from pairs of face parts reveals a high potential to define effective combinations of classifiers.

1 Introduction

Faces are the core of human interaction. In particular, faces present evident sexual information which makes gender recognition a fast, efficient cognitive process. Even in cases in which exogenous and cultural features such as hair and make-up are removed, adults are able to correctly recognize gender in about 96% of cases. In recent years, a number of papers have paid attention to the problem of automatic gender recognition from face image inspection [1–4]. Possible applications can be imagined in access control, in the demographic description of a population, and so forth.

In [1], 200 face data from a particular dataset, 100 for each gender, were represented in terms of face surface normals. A two-component mixture model was trained by EM on selected features, reaching an overall 95.5% of correct classifications. A wider experimental study was performed in [2], where a more standard dataset was used. Given a face image, a large number of local regions feed a feature extraction method based on multi-scale filter banks [5]. Classification was defined from a distance-based criterion considering, first, all features together and, second, the features separated into two subsets depending on the

zone in which they were located, i.e. internal ones extracted from eyes, nose and mouth, and external features derived from head, ears and face contour. Gender recognition was carried out both in a highly controlled environment and in an uncontrolled context, as far as illumination and background conditions are concerned. The best accuracies were 96.77% and 91.72% for controlled and uncontrolled environments, respectively, using all features together in both situations.

In [2], the classifier computes the contribution of local regions to class decisions, but the regions are not associated with specific face parts (eye, nose, mouth, etc.). Therefore, no conclusions about the contribution of face parts are possible. Why could it be interesting to evaluate the effectiveness of using face parts in gender discrimination? A first answer could be related to classification under partial occlusions of the face. Once a visible part, for example an eye, is identified and used to predict a gender, the knowledge of its effectiveness can be taken as the reliability of the decision. An extended approach could evaluate the joint efficacy of two or more visible face parts, which can be used in a joint prediction of a gender.

Two recent papers [3, 4] have studied the importance of face parts in gender recognition. In [3], the differentiation capabilities of full face, jaw, lips/mouth, nose and eyes were evaluated. These regions were manually clipped, represented by an appearance-based method, and classified using linear discriminant analysis. The best recognition rates were 93.7% and 89.8% by the full face and jaw, respectively, while the worst accuracy rates were lower than 80% and correspond to nose and eyes. These results are certainly contrary to general intuition, according to which eyes, nose or mouth seem to be more relevant than the jaw to discriminate between women and men. Apart from subjective judgments, the generalization of these results is limited if we take into account the fact that the database consists only of expressionless Asian faces.

Other closely related work is [4], where the roles of the full face, eyes and mouth in gender classification were compared. The subimages of internal features (eyes and mouth) were extracted and their dimensionality reduced using PCA, CCA and SOM. Classification was performed by a SVM [6] with a RBF kernel, and results were obtained over an ad hoc set of 400 face images selected from three well-known databases: FERET [7], AR [8] and BioID [9]. The best gender recognition rate was achieved from a joint representation of 759 features obtained from the full face, eyes and mouth. When individual parts were considered, the best accuracies were 85.5% and 81.25% for eyes and mouth, respectively. In this work, where more general face data (w.r.t. [3]) and three different data transformations were available, nose and jaw were not evaluated and hair was not removed from full faces. The authors admit that hair had a dominating (and negative) effect on gender classification based on full faces.

Considering that results about the discriminant capabilities of face parts obtained in [3, 4] depend on specific components (dataset, classifier, face part description), care should be taken when interpreting their conclusions. For example, the relation between gender recognition rates from eyes and mouth is

inverted in these two papers, eyes being more discriminant than mouth in [4], and mouth more accurate than eyes in [3].

This paper goes beyond these two previous related works, as regards the number of face parts and the diversity of the experimental design. In particular, gender recognition is performed from descriptions of single parts like eyes (including eyebrows), nose, mouth, chin and right eye, and from descriptions of global regions like internal faces (eyes, nose, mouth, chin), external faces (hair, ears, contour) and full faces. Experiments were defined for two standard datasets, FERET [7] and XM2VTS [10], and several classifiers: SVM [6], {1,5,10}-NN, Quadratic Bayes Normal and Parzen. Apart from a comparison of the importance of different face parts, we also discuss some conclusions about the complementarity between pairs of parts which can be useful for combining classifiers.

2 Methodology

2.1 From Images to Vectors

Given a frontal image of a face and the coordinates of the two eyes, this section explains how 8 subimages containing the eyes, nose, mouth, chin, right eye, internal face, external face and the full face are defined from expected proportions of an aesthetic face and transformed into data vectors.

Before this data extraction procedure, images were transformed into gray format and preprocessed by means of histogram equalization in order to improve contrast and make lighting conditions more uniform.

The process of subimage extraction is based on an empirical rule about the ideal balance of a human face sketched by Leonardo da Vinci [11, 12]. Leonardo stated that perfect facial harmony exists when the face can be divided vertically into three equal sections whose boundaries match with the hairline, the eyebrows, the bottom of the nose and the chin, and when the face can be partitioned horizontally into five sections that approximate the width of one eye. In this paper, the horizontal division was carried out as defined in the previous rule, but vertical division was performed following a particular layout, in which some key face features were centered in their corresponding regions (see Fig. 1(a)). The division points were automatically computed from the knowledge of eye coordinates.

The rectangular regions of the proposed grid can fully enclose eyes, but other important parts like mouth, nose, and chin are bounded in only one direction. For example, the nose is bounded in the vertical direction, while the mouth is only bounded horizontally. To fully describe these features, new zones, which are referred to as subimages, are created by joining adjacent (parts of) rectangular regions (see Fig. 1(b)). The resulting subimages are then scaled down to low-resolution regions, where new pixels are computed by averaging the original ones. This process was conducted by reducing each rectangular region of the grid to a new 6x6 rectangular window. This paper has not explored other sizes which could provide even more accurate descriptions and better results. Figure 2

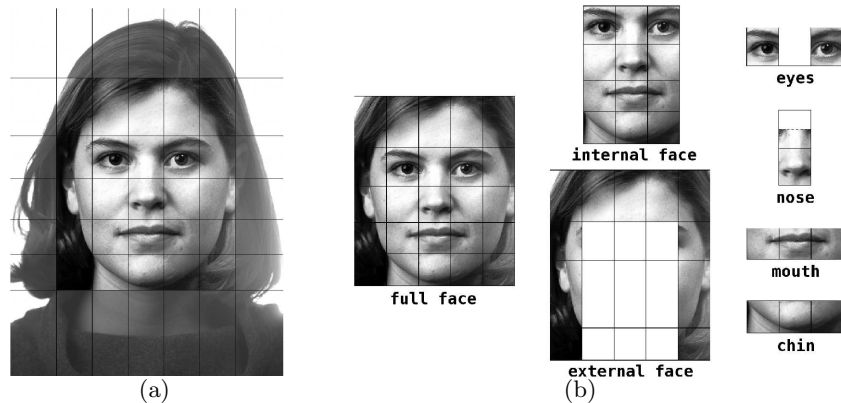


Fig. 1. Images involved in the vectors extraction process. (a) Image with the face preprocessed and the grid. (b) Subimages which are transformed into vectors (the right eye's subimage is extracted from the eyes subimage).

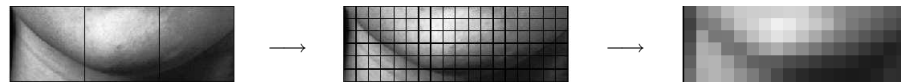


Fig. 2. Example of scaling chin's subimage down to a low-resolution representation.

illustrates this process. The new reduced subimages are then transformed into linear vectors.

2.2 From Vectors to Gender Recognition

Given a dataset of face images with gender labels, the previous procedure can generate 8 new labeled datasets, each composed of all the vectors that describe the same part in all the faces (including the full face). From now on, this task can be managed as a standard pattern recognition problem. In fact, there are 8 simple related tasks with the same number of instances and the same set of labels, which can be handled as individual problems or used in a combined approach where diversity would be supported by different data representation spaces.

In this paper only individual classifiers were considered, although possible ensembles have been explored. In order to reduce dimensionality and to boost data information, PCA was applied to data vectors before classification. The following section gives details about the learning and recognition processes.

3 Experiments

Experiments involved two standard datasets of face images, FERET [7] and XM2VTS [10]. Only faces in facing the front without glasses were used, because

Table 1. Number of attributes of FERET datasets before and after applying PCA.

	eyes	nose	mouth	chin	right eye	internal face	external face	full face
before PCA	72	54	108	108	36	432	468	900
after PCA	38	29	48	36	21	62	59	101

Table 2. Gender recognition rates on FERET datasets.

	eyes	nose	mouth	chin	right eye	internal face	external face	full face
SVM	85.47	86.36	81.61	81.56	81.51	92.37	87.71	95.21
1-NN	77.51	78.39	75.37	76.76	73.04	83.28	84.22	86.54
5-NN	82.07	82.49	79.79	78.53	79.65	87.48	86.17	87.89
10-NN	82.91	83.38	80.68	79.70	80.12	88.03	86.31	87.62
QDC	84.03	81.23	79.46	81.14	82.21	89.11	89.25	92.37
Parzen	80.35	81.88	79.74	78.67	76.81	85.38	85.98	87.44

the presence of glasses could strongly distort the effectiveness of eyes as a means of gender recognition. From each image set, 8 related vector datasets were inferred (see Sect. 2.1), including descriptions of eyes, nose, mouth, chin, right eye, internal face, external face and full face. These datasets were transformed by PCA, where new attributes were selected to explain 99% and 95% of the total variance for the 5 single parts and the 3 global parts, respectively. The figure of 95% for global parts allows greater reductions in datasets to be made with more original attributes. Several learning algorithms were taken into account, i.e. SVM [6] with a linear polynomial kernel, {1,5,10}-Nearest Neighbor rules, the Quadratic Bayes Normal classifier (QDC in [13]) and the Parzen classifier. Classification results were computed from the implementations available in the PRTools Matlab package [13] with their default parameter values. Errors were estimated by a 5-fold cross-validation technique, where all face images of the same person were included in the same subset to avoid contamination effects between training and test partitions.

3.1 Results on FERET Database

FERET [7] is a dataset of face images acquired in a controlled environment, but with no restrictions as to age, race, or face expressions. It is composed of several images of the same person, which exist in three different sizes divided into three collections. Experiments in this section involved 2147 medium-sized images of 256 x 384 pixels from 834 subjects, separated into 842 female faces and 1305 male faces. As was indicated in Sect. 2.1, subimages were scaled down to low-resolution representations. Table 1 shows the number of appearance-based attributes that represent each face part, and the final dimensionality after applying PCA.

The gender recognition results on the FERET database are shown in Tab. 2 and Fig. 3. The first point to be considered is related to the high discriminant power of individual face parts: most of them produce classification rates above

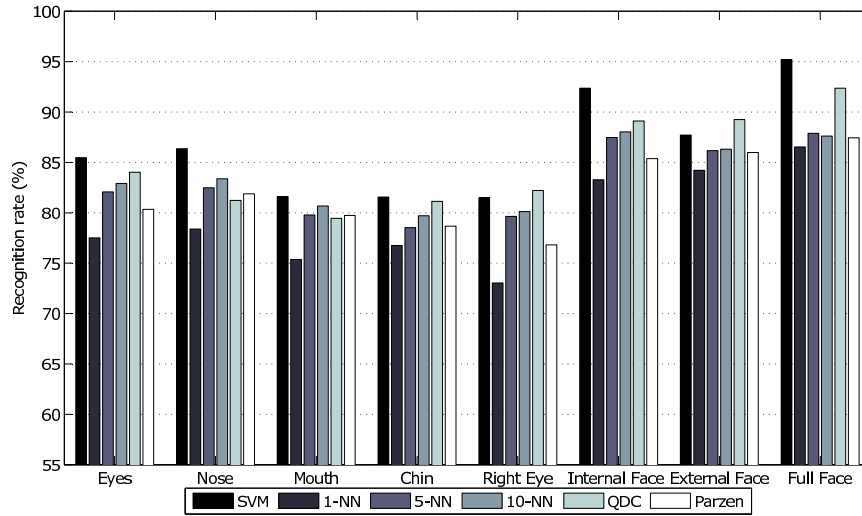


Fig. 3. Gender recognition rates on FERET. The best result was obtained by the SVM trained from full faces. Nose and eyes were the most discriminat individual parts.

80%, mainly for the SVM, QDC, 10-NN and Parzen classifiers. With regard to the relative importance of each of them, nose was the most relevant part in identifying sex for all classifiers except QDC, for which learning from eyes led to a better result. Moreover, these two parts were more discriminant than mouth and chin for all classifiers.

As expected, gender recognition based on global parts of the face was more accurate than that based on individual parts. In particular, when SVM and QDC learn from internal, external and full face, rates were very close to or higher than 90%, and in the case of SVM with a full face dataset the rate was above 95%. It is also interesting to remark that external face by itself, dominated by hair, was almost as discriminant as the internal part of the face (see Fig. 1(b)). Bearing in mind that the full face is the integration of the internal and the external parts, the effective contribution (in terms of new information) of each part with respect to the other was very small (no greater than 3%).

3.2 Results on XM2VTS Database

The XM2VTS [10] database was created at the Centre for Vision, Speech and Signal Processing at the University of Surrey. In this experiment, 1378 images of 720x576 pixels from 203 subjects were divided into 732 female faces and 646 male faces. The numbers of attributes of the face parts in the datasets before and after applying PCA are shown in Tab. 3.

The recognition results on the XM2VTS database, shown in Tab. 4 and in Fig. 4, are clearly lower than those obtained with FERET. In this experiment,

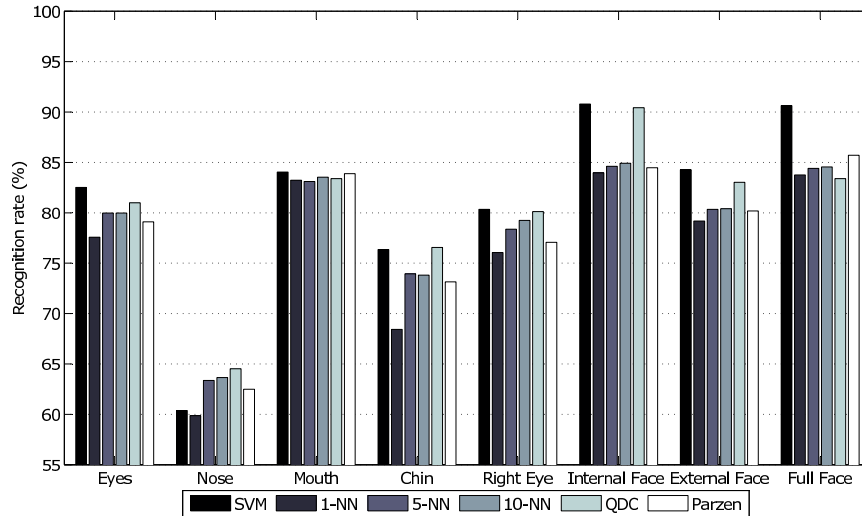


Fig. 4. Gender recognition rates on XM2VTS. The best result was obtained by SVM trained with internal faces. Mouth and eyes were the most discriminat individual parts.

not all individual parts turned out to be as effective for recognizing gender as in FERET. Taking into account their relative importance, mouth and eyes (including the single right eye) were the most discriminant parts for all classifiers. Good results were also obtained for the chin dataset, while rates derived from the nose were surprisingly low.

Although global parts of the face are, in general, more effective than individual parts for gender classification, there are important singularities that can be highlighted. Firstly, the differences between global parts and the best individual part (the mouth) is negligible for 4 (similar) classifiers: {1,5,10}-NN and Parzen. Secondly, the internal part of the face appears to be as discriminant as the full face, which suggests that the external part does not contribute any new information. In fact, the external part produces lower recognition rates than the mouth.

3.3 Discussion of Results

Eyes gave rise to good classifiers in both databases, while mouth and chin were also effective. The nose, in spite of being the best individual part in FERET, pro-

Table 3. Number of attributes of XM2VTS datasets before and after applying PCA.

	eyes	nose	mouth	chin	right eye	internal face	external face	full face
before PCA	72	54	108	108	36	432	468	900
after PCA	42	33	54	48	23	74	76	125

Table 4. Gender recognition rates on XM2VTS datasets.

	eyes	nose	mouth	chin	right eye	internal face	external face	full face
SVM	82.52	60.38	84.04	76.35	80.34	90.79	84.26	90.64
1-NN	77.58	59.87	83.24	68.44	76.06	83.97	79.18	83.75
5-NN	79.98	63.36	83.10	73.95	78.38	84.62	80.34	84.40
10-NN	79.98	63.65	83.53	73.81	79.25	84.91	80.41	84.55
QDC	80.99	64.52	83.39	76.56	80.12	90.43	83.02	83.39
Parzen	79.10	62.49	83.89	73.15	77.07	84.47	80.19	85.71

duced very poor results in XM2VTS. This could be explained by the prominence of the nose, which causes a higher exposure to changes in illumination and could make the description of the nose more unstable. Apart from this specific hypothesis, the significant difference between the FERET and XM2VTS databases could have also contributed to the poorer classification results in XM2VTS. More specifically, the number of subjects in FERET is about 4 times greater than the number in XM2VTS, while the average number of frontal images per subject in FERET is about half the index in XM2VTS. Therefore, a wider representation of the problem is obtained in the case of FERET. With respect to the influence of classifiers on the relevance of face parts, results show a high correlation among their behaviors and a strong degree of independence from them. It is worth noting that the external part of the face (hair, ear, contour) can also be useful to predict gender, since there is a high correlation between gender and traditional cultural patterns (length of hair, use of earrings).

3.4 On the Evaluation of Complementarity of Face Parts

The fact that a subject’s face image is simultaneously described in different Euclidean spaces, coming from different face parts, has triggered research into the possibility of creating ensembles based on the diversity of such spaces.

Table 5 shows the SVM disagreement when learning from each pair of face parts on FERET instances. Each number represents the percentage of instances which are successful cases for the row classifier and error cases for the column classifier. In other words, each percentage is the potential improvement that the row classifier can afford to the column rule. Obviously, the performances of classifiers with lower rates of success (those based on mouth and chin) have a greater chance of being enhanced (about 12-13%) than other rules with better accuracies, for example, the one based on the full face, with about 1-2%.

Some particular situations are now reviewed. For example, the performance of the eyes-based SVM could be potentially raised from 85.47% to more than 95% by any of the other classification rules, because any of them could correct about 10% of the errors of the former. In the same way, the performance of SVM trained from full faces could (possibly) be improved from 95.21% up to more than 97% by the eyes-based SVM. This analysis reveals the existence of significant complementarities between pairs of face parts, which suggests that

Table 5. Results of SVM disagreement when learning from each pair of face parts on FERET instances. Each number represents the percentage of instances which are success cases for the row classifier and error cases for the column classifier.

	eyes	nose	mouth	chin	internal face	external face	full face
eyes	-	8.8	13.97	14.01	3.35	8.01	2.46
nose	9.68	-	11.73	12.90	3.07	7.49	1.63
mouth	10.10	6.98	-	8.57	2.65	6.28	1.76
chin	10.10	8.10	8.52	-	3.16	6.42	1.95
internal	10.24	9.08	13.41	13.97	-	9.08	2.14
external	10.24	8.84	12.38	12.57	4.42	-	1.02
full	12.20	10.47	15.37	15.60	4.98	8.52	-

effective and robust ensembles can be proposed. The problem of how to design them is beyond the scope of this paper and could be studied in future work.

4 Conclusions

This paper has evaluated the relevance of face parts in gender recognition, and considered its potential usefulness in classification under partial occlusions. Based on an empirical rule about the ideal balance of an aesthetic face proposed by Leonardo da Vinci, subimages containing eyes, nose, mouth, chin, right eye, internal face, external face, and full face are extracted and transformed into appearance-based data vectors. Experiments are based on two known databases of face images and five classification rules.

This paper goes beyond other related works in terms of the numbers of face parts, databases and classifiers employed. Consequently, more general and robust conclusions have been obtained.

As the experiments have proved, individual face parts include enough information to be able to discriminate between genders with success rates above 80%. When their joint contribution was considered by means of the full face or by just its internal part, improvements in recognition rates of over 95% were obtained. This result is similar to those reported in general papers of gender recognition.

Experiments were also designed to evaluate the dependence of these results on the database and the classifier that were used. A high correlation among classifiers on measuring the relevance of face parts was obtained, but results were strongly dependent on the database. The significant differences between the two databases were transferred to results.

The possibility of creating ensembles based on diversity of the different representations of a same face image was also studied. For each combination of two face parts, we computed the percentage of cases in which SVM trained on them did not coincide. These statistics showed a high potential to define robust mixtures of classifiers. This task, together with gender recognition under occlusions, are two interesting future lines of work.

5 Acknowledgments

This work was partially funded by projects CSD2007-00018 and DPI2006-15542 from the Spanish Ministry of Science and Education, and P1-1B2007-60 from the Fundació Caixa Castelló-Bancaixa.

References

1. Wu, J., Smith, W., Hancock, E.: Learning mixture models for gender classification based on facial surface normals. In: LNCS: Proc. of 3rd IBPRIA. Volume 4477., Girona, Spain, Springer (2007) 39–46
2. Lapedriza, A., Marín-Jiménez, M., Vitrià, J.: Gender recognition in non controlled environments. In: Proc. of 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, IEEE (2006)
3. Kawano, T., Kato, K., Yamamoto, K.: A comparison of the gender differentiation capability between facial parts. In: Proc. of 17th International Conference on Pattern Recognition (ICPR'06), Cambridge, U.K., IEEE (2004)
4. Buchala, S., Davey, N., Frank, R., Gale, T., Loomes, M., Kanargard, W.: Gender classification of faces images: The role of global and feature-based information. In: LNCS: Proc. of 11th Int. Conf. on Neural Information Processing. Volume 3316., Calcutta, India, Springer (2004) 763–768
5. Marín-Jiménez, M., Pérez de la Blanca, N.: Empirical study of multi-scale filter banks for object categorization. In: Proc. of 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, IEEE (2006)
6. Cortes, C., Vapnik, V.: Support-vector networks. *Machine Learning* **20**(3) (1995) 273–297
7. Phillips, H., Moon, P., Rizvi, S.: The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(10) (2000)
8. Martinez, A., Benavente, R.: The AR face database. Technical Report 24, CVC (June 1998)
9. Jesorsky, O., Kirchberg, K., Frischholz, R.: Robust face detection using the hausdorff distance. In Bigun, J., Smeraldi, F., eds.: *Audio and Video based Person Authentication: Proc. of AVBPA 2001*, Springer (2001) 90–95
10. Messer, K., Matas, J., Kittler, J., Lüttin, J., Maitre, G.: XM2VTSDB: The extended M2VTS database. In: *Audio and Video-based Biometric Person Authentication, AVBPA'99*. (1999) 72–77
11. Nainia, F.B., Mossb, J.P., Gilc, D.S.: The enigma of facial beauty: Esthetics, proportions, deformity, and controversy. *American Journal of Orthodontics and Dentofacial Orthopedics* **130**(3) (2006) 277–282
12. Oguz, O.: The proportion of the face in younger adults using the thumb rule of Leonardo da Vinci. *Surgical and Radiologic Anatomy* **18**(2) (1996) 111–114
13. Duin, R., Juszczak, P., Paclik, P., Pekalska, E., de Ridder, D., D.M. Tax, J.: *PRTools4, A Matlab Toolbox for Pattern Recognition*. Delft University of Technology, The Netherlands. 4.0 edn. (2004)